What is claimed is:

1. A computer-readable medium storing a computer program for use in conjunction with a web browser client program to rate a web page relative to a selected characteristic, the program comprising:

first means for identifying natural language textual portions of the web page and forming a list of words that appear in the identified natural language textual portions of the web page;

a database of predetermined words that are associated with the selected characteristic;

second means for querying the database to determine which of the list of words has a match in the database;

third means for acquiring a corresponding weight from the database for each such word having a match in the database so as to form a weighted set of terms; and

fourth means for calculating a rating for the web page responsive to the weighted set of terms, the calculating means including means for determining and taking into account a total number of natural language words that appear in the identified natural language textual portions of the web page.

2. A computer-readable medium storing a computer program for use in conjunction with a web browser client program to rate a web page according to claim 1 wherein the selected characteristic is pornographic content; and

the database includes a predetermined list of words and phrases that are associated with web pages having pornographic content.

3. A computer-readable medium storing a computer program for use in conjunction with a web browser client

program to rate a web page according to claim 1 wherein the calculating means includes means for summing the weights for each word having a match in the database.

4. A computer-readable medium storing a computer program for use in conjunction with a web browser client program to rate a web page according to claim 1 and further comprising means for storing a predetermined threshold rating, and means for comparing the calculated rating to the threshold rating to determine whether the web page likely has the selected characteristic.

5. A method of analyzing content of a digital data set, the method comprising the steps of:

identifying natural language textual portions of the data set;

forming a word list including all natural language words that appear in the textual portions of the data set;

for each word in the word list, querying a preexisting database of selected words to determine whether or not a match exists in the database;

for each word having a match in the database, reading a corresponding weight from the database so as to form a weighted set of terms; and

calculating a rating for the data set responsive to the weighted set of terms.

6. A method according to claim 5 wherein the data set comprises a web page.

7. A method according to claim 5 wherein the data set comprises a web page and further comprising:

identifying meta-content in the web page; and

including words from the meta-content of the web page in the word list so that the meta-content is taken into account in calculating the rating for the web page.

8. A method according to claim 5 wherein said calculating step includes:

summing the weighted set of terms together to form a sum;

multiplying the sum by a predetermined modifier to scale the sum;

determining a total number of words on the web page; and

dividing the scaled sum by the total number of words on the data set to form the rating.

9. A method according to claim 5 wherein the data set comprises an email.

10. A method of building a target attribute set for use in analyzing content of a digital data set, the method comprising the steps of:

acquiring a plurality of sample data sets for use as training data sets;

designating each of the training data sets as "yes" or "no" with respect to a predetermined content characteristic;

parsing through the content of all of the training data sets to form a list of regular expressions that appear in the training data sets;

forming data reflecting a frequency of occurrence of each regular expression in the training data sets;

analyzing the frequency of occurrence data in view of the "yes" or "no" designation of each data set, to identify and select a set of regular expressions that are indicative

of either a "yes" designation or a "no" designation of a data set with respect to the predetermined characteristic; and

storing the selected set of regular expressions to form a target attribute set based on the downloaded training pages, whereby the target attribute set provides a set of regular expressions that are useful in discriminating data set content relative to the predetermined content characteristic.

11. A method of building a target attribute set according to claim 10 wherein the digital data set comprises a web page; the sample data sets comprise a plurality of web pages downloaded from the world-wide web; and the predetermined content characteristic is pornographic content.

12. A method of building a target attribute set according to claim 11 further comprising selecting a number of web pages on the order of 1,000 web pages for use as training data sets.

13. A method of building a target attribute set according to claim 10 wherein the digital data set comprises an email message; the sample data sets comprise a plurality of email messages; and the predetermined content characteristic is pornographic content.

14. A method of assigning weights to a list of regular expressions for use in analyzing content of a digital data set, the method comprising:

providing a predetermined target attribute set associated with a predetermined group of training data sets, the target attribute set including a list of regular

expressions that are deemed useful for discriminating data set content relative to a predetermined content characteristic;

assigning an initial weight to each of the regular expressions in the target attribute set, thereby forming a weight database;

designating each of the group of training data sets as either "yes" or "no" relative to whether it exhibits the predetermined content characteristic;

examining one of the group of training data sets to identify all regular expressions within the data set that also appear in the target attribute set, thereby forming a match list for said data set;

in a neural network system, rating the examined data set using the weightings in the weight database;

comparing the rating of the examined data set to the corresponding "yes" or "no" designation to form a first error term;

repeating said examining, rating and comparing steps for each of the remaining data sets in the group of training data sets to form additional error terms; and

adjusting the weights in the weight database in response to the first and the additional error terms.

15. A method of assigning weights according to claim 14 wherein the predetermined content characteristic is pornography.

16. A method for controlling access to potentially offensive or harmful web pages comprising the steps of:

in conjunction with a web browser client program executing on a digital computer, examining a downloaded web page before the web page is displayed to the user; said examining step including analyzing the web page natural

language content relative to a predetermined database of words to form a rating, the database including words previously associated with potentially offensive or harmful web pages, and the database further including a relative weighting associated with each word in the database for use in forming the rating;

comparing the rating of the downloaded web page to a predetermined threshold rating; and

if the rating indicates that the downloaded web page is more likely to be offensive or harmful than a web page having the threshold rating, blocking the downloaded web page from being displayed to the user.

17. A method according to claim 16 further comprising:

if the downloaded web page is blocked, displaying an alternative web page to the user.

18. A method according to claim 17 wherein said displaying an alternative web page includes generating or selecting the alternative web page responsive to a predetermined categorization of the user.

19. A method according to claim 17 wherein the alternative web page includes an indication of the reason that the downloaded web page was blocked.

20. A method according to claim 16 wherein the alternative web page includes one or more links to other web pages selected as age-appropriate in view of a predetermined categorization of the user.

21. A computer-readable medium storing a web search engine server program, the program comprising:

a data acquisition component for acquiring meta-content from target web sites into an internal database; and

an inquiry component for selecting and presenting meta-content from the internal database in response to an end-user request;

the data acquisition component including an analysis component that analyzes the content of web pages corresponding to the meta-content stored in the internal database, and returns a rating for each such web page; and

means for adding said returned ratings into the internal database as additional meta-content in association with the corresponding web pages.

22. A computer-readable medium storing a web search engine server program according to claim 21, the analysis component including:

first means for identifying natural language textual portions of the web page and forming a list of words that appear in the identified natural language textual portions of the web page;

a second internal database of predetermined words that are associated with the selected characteristic;

second means for querying the second internal database to determine which of the list of words has a match in the database;

third means for acquiring a corresponding weight from the second internal database for each such word having a match in the second internal database so as to form a weighted set of terms; and

fourth means for calculating a rating for the web page responsive to the weighted set of terms, the calculating means including means for determining and taking into account a total number of natural language words that

appear in the identified natural language textual portions of the web page.

23. A computer-readable medium storing a web search engine server program according to claim 21, and further comprising means for including the additional meta-content in said presenting meta-content from the internal database in response to an end-user request.

24. A computer-readable medium storing a web search engine server program according to claim 21, and further comprising means for modifying the meta-content results presented in response to an end-user request based upon the said ratings.